Supplementary material for 'Optimistic bounds for multi-output prediction'

A. The self-bounding Lipschitz condition

A.1. Proof of Lemma 1

The proof of Lemma 1 starts with the following lemma.

Lemma 7. *Suppose that $\varphi : \mathbb{R} \to [0, \infty)$ is a non-negative differentiable function satisfying:*

1. *The derivative $\varphi'(t)$ is non-negative on $[0, \infty)$;*
2. $\forall t_0, t_1 > 0$, $|\varphi'(t_1) - \varphi'(t_0)| \leq \left(\frac{\lambda}{2}\right)^{\frac{1}{1-\theta}} \cdot |t_1 - t_0|^{\frac{\theta}{1-\theta}}$.

Then $\forall t > 0$, $\varphi'(t) \leq \lambda \cdot \varphi(t)^\theta$. Moreover, for all $t > 0$, $\varphi(t) - \varphi(0) \leq \lambda \cdot \varphi(t)^\theta \cdot t$.

Proof. Fix $t > 0$ and take $\Delta = 2\lambda^{-\frac{1}{\theta}} \cdot \varphi'(t)^{\frac{1-\theta}{\theta}}$, which is positive by the first condition. By the non-negativity of φ and the mean value theorem there exists some $s \in (t - \Delta, t)$

$$\begin{aligned}
0 \leq \varphi(t - \Delta) &\leq \varphi(t) - \varphi'(s) \cdot \Delta \\
&\leq \varphi(t) - \varphi'(t) \cdot \Delta + |\varphi'(s) - \varphi'(t)| \cdot \Delta \\
&\leq \varphi(t) - \varphi'(t) \cdot \Delta + \left((\lambda/2)^{\frac{1}{1-\theta}} \cdot \Delta^{\frac{\theta}{1-\theta}}\right) \cdot \Delta \\
&\leq \varphi(t) - \varphi'(t) \cdot \Delta + (\lambda \cdot \Delta/2)^{\frac{1}{1-\theta}} \\
&\leq \varphi(t) - 2(\varphi'(t)/\lambda)^{\frac{1}{\theta}} + (\varphi'(t)/\lambda)^{\frac{1}{\theta}} \\
&= \varphi(t) - (\varphi'(t)/\lambda)^{\frac{1}{\theta}},
\end{aligned}$$

where the fourth inequality follows from the second condition. Rearranging completes the proof of the first part of the lemma.

To prove the second part of the lemma we apply the mean value theorem combined with the first part of the lemma to obtain for some $s \in (0, t)$,

$$\varphi(t) - \varphi(0) = \varphi'(s) \cdot t \leq \left(\lambda \cdot \varphi(s)^\theta\right) \cdot t \leq \lambda \cdot \varphi(t)^\theta \cdot t,$$

where we used the non-negativity of φ' on $[0, \infty)$ to ensure that $\varphi(s) \leq \varphi(t)$. This completes the proof of the lemma. □

Proof of Lemma 1. Take $u, v \in \mathcal{V}$ and $y \in \mathcal{Y}$. Without loss of generality we assume that $\mathcal{L}(u, y) \leq \mathcal{L}(v, y)$ and let $\varphi_{u,y}$ be a function satisfying the conditions specified in the statement of the lemma. By combining the first two conditions with Lemma 7 we see that $\varphi_{u,y}(t) - \varphi_{u,y}(0) \leq \lambda \cdot \varphi_{u,y}(t)^\theta \cdot t$. Hence, by dividing through by $\varphi_{u,y}(t)^\theta$ and applying $\mathcal{L}(v, y) \leq \varphi_{u,y}(t)$ twice we have,

$$\begin{aligned}
\mathcal{L}(v, y)^{1-\theta} - \lambda \cdot t &\leq \varphi_{u,y}(t)^{1-\theta} - \lambda \cdot t \\
&\leq \varphi_{u,y}(0) \cdot \varphi_{u,y}(t)^{-\theta} \\
&\leq \varphi_{u,y}(0) \cdot \mathcal{L}(v, y)^{-\theta} \\
&= \mathcal{L}(u, y) \cdot \mathcal{L}(v, y)^{-\theta}.
\end{aligned}$$

Multiplying by $\mathcal{L}(v, y)^\theta$ and rearranging we have $\mathcal{L}(v, y) - \mathcal{L}(u, y) \leq \lambda \cdot \mathcal{L}(v, y)^\theta$. Since $\mathcal{L}(u, y) \leq \mathcal{L}(v, y)$ this completes the proof of the lemma. □

A.2. Proof of Lemma 2

Proof of Lemma 2. Take $u, v \in \mathcal{V}$ and $y \in \mathcal{Y}$. Without loss of generality we assume that $\tilde{\mathcal{L}}(u, y) \leq \tilde{\mathcal{L}}(v, y)$, so it suffices to show that

$$\tilde{\mathcal{L}}(v, y) - \tilde{\mathcal{L}}(u, y) \leq \lambda \cdot \tilde{\mathcal{L}}(v, y)^\theta \cdot \|u - v\|_\infty. \tag{3}$$

If $\mathcal{L}(u,y) \geq b$ then $\tilde{\mathcal{L}}(v,y) = \tilde{\mathcal{L}}(u,y) = b$, so (3) clearly holds. Thus, we can assume $\mathcal{L}(u,y) < b$, so $\tilde{\mathcal{L}}(u,y) = \mathcal{L}(u,y)$. By the (λ, θ) self-bounding Lipschitz condition for $\mathcal{L}$ we have

$$\begin{aligned}\mathcal{L}(v,y) - \tilde{\mathcal{L}}(u,y) &= \mathcal{L}(v,y) - \mathcal{L}(u,y)\\ &\leq \lambda \cdot \mathcal{L}(v,y)^{\theta} \cdot \|u-v\|_{\infty}.\end{aligned}$$

Equivalently, we have

$$\mathcal{L}(v,y)^{1-\theta} - \lambda \cdot \|u-v\|_{\infty} \leq \tilde{\mathcal{L}}(u,y) \cdot \mathcal{L}(v,y)^{-\theta}.$$

Since $\tilde{\mathcal{L}}(v,y) \leq \mathcal{L}(v,y)$, we deduce

$$\tilde{\mathcal{L}}(v,y)^{1-\theta} - \lambda \cdot \|u-v\|_{\infty} \leq \tilde{\mathcal{L}}(u,y) \cdot \tilde{\mathcal{L}}(v,y)^{-\theta}.$$

Rearranging gives (3) and completes the proof of the lemma. □

A.3. Proof of Proposition 2

The following result shows an example application of Lemma 1. We may verify the self-bounding Lipschitz condition for other loss functions in a similar manner.

Proposition 2. *Take $\mathcal{Y} = [q]$ and define the multinomial logistic loss $\mathcal{L} : \mathcal{V} \times \mathcal{Y} \to [0, \infty)$ is defined by*

$$\mathcal{L}(u,y) = \log(\sum_{j\in[q]} \exp(u_j - u_y)),$$

*where $u = (u_j)_{j\in[q]}$ and $y \in [q]$. It follows that $\mathcal{L}$ is (λ, θ)-*self-bounding Lipschitz *with $\lambda = 1$ and $\theta = 1/2$.*

The proof of Proposition 2 requires the following elementary lemma.

Lemma 8. *Given any $A > 0$ the function $\varphi : \mathbb{R} \to (0, \infty)$ defined by $\varphi(t) = \log(1 + A \cdot \exp(2t))$ is differentiable $\varphi'(t_0) > 0$ and and $|\varphi'(t_0) - \varphi'(t_1)| \leq |t_1 - t_0|$ for all $t_0, t_1 \in \mathbb{R}$.*

Proof. We begin by computing the first three derivatives,

$$\begin{aligned}\varphi'(t) &= \frac{2A \cdot \exp(2t)}{1 + A \cdot \exp(2t)}\\ \varphi''(t) &= \frac{4A \cdot \exp(2t)}{(1 + A \cdot \exp(2t))^2}\\ \varphi'''(t) &= \frac{8A \cdot \exp(2t)}{(1 + A \cdot \exp(2t))^3} \cdot (1 - A \cdot \exp(2t)).\end{aligned}$$

Clearly we have $\varphi'(t), \varphi''(t) > 0$ for all $t \in \mathbb{R}$. Moreover, by inspecting the third derivative we see that φ'' has a unique maximum where $A \cdot \exp(2t) = 1$. This implies that φ is twice differentiable with $|\varphi''(t)| \leq 1/4$ for all $t \in \mathbb{R}$. By the mean value theorem this yields $|\varphi'(t_0) - \varphi'(t_1)| \leq |t_1 - t_0|$ for all $t_0, t_1 \in \mathbb{R}$. □

Proof of Proposition 2. To complete the proof we $A_{u,y} := \sum_{j\in[q]\setminus\{y\}} \exp(u_j - u_y)$ and define $\varphi_{u,y}(t) := \log(1 + A_{u,y} \cdot \exp(2t))$. We can apply Lemma 8 to confirm that $\varphi_{u,y}$ satisfies the conditions of Lemma 1. Hence, the conclusion of Proposition 2 follows from Lemma 1. □

B. Proof of Theorem 2

For completeness here we give a proof of Theorem 2, which may be viewed as a mild generalisation of Theorem 6 from (Lei et al., 2019). We use the following well known result.

Theorem 8 ((Bartlett & Mendelson, 2002)). *Suppose we have a measurable space $\mathcal{Z}$ and a function class $\mathcal{G} \subseteq \mathcal{M}(\mathcal{Z}, [0, b])$. For each $\boldsymbol{z} \in \mathcal{Z}^n$ and $g \in \mathcal{G}$ we let $\hat{\mathbb{E}}_{\boldsymbol{z}}(g) = n^{-1} \cdot \sum_{i \in [n]} g(z_i)$. Suppose that Z is a random variable with distribution P is a distribution on $\mathcal{Z}$ and let $\mathcal{D} = \{Z_i\}_{i \in [n]} \in \mathcal{Z}^n$ be an i.i.d. which each $Z_i \sim P$, an independent copy of Z. For any $\delta \in (0, 1)$, the following holds with probability at least $1 - \delta$, for all $g \in \mathcal{G}$,*

$$\left| \mathbb{E}_Z(g) - \hat{\mathbb{E}}_{\boldsymbol{z}}(g) \right| \leq 2\mathbb{E}_{\mathcal{D}} \left[\hat{\mathfrak{R}}_{\mathcal{D}} \left(\mathcal{G} \right) \right] + \sqrt{\frac{\log(2/\delta)}{2n}}.$$

Proof of Theorem 2. With the correspondence introduced in the proof of Theorem 1, Theorem 8 implies that with probability at least $1 - \delta$ over a sample $\mathcal{D} = \{(X_i, Y_i)\}_{i \in [n]}$ with $(X_i, Y_i) \sim P$ the following holds for all $f \in \mathcal{F}$,

$$\left| \mathcal{E}_{\mathcal{L}}(f, P) - \hat{\mathcal{E}}_{\mathcal{L}}(f, \mathcal{D}) \right| \leq 2\mathbb{E}_{\mathcal{D}} \left[\hat{\mathfrak{R}}_{\mathcal{D}} \left(\mathcal{L} \circ \mathcal{F} \right) \right] + \sqrt{\frac{\log(2/\delta)}{2n}}.$$

Hence, the result follows from Proposition 1 by taking $r = b$ and $\theta = 0$. $\square$